

Tour d'horizon des 11 technologies prometteuses du Big Data

Sous-titrée *"Big Data Is Critical Technology For Insights-Driven Businesses"*, l'enquête TechRadar: Big Data, Q1 2016 du cabinet d'études Forrester Research explique et évalue le potentiel de ces 22 technologies Big Data. Ces dernières pouvant être déployées sur site, sous forme de service cloud, ou en mode hybride.

«Les technologies Big Data décrites dans cette étude font partie des trois catégories intervenant dans le cycle de vie de la donnée : gérer en prenant en compte l'évolutivité, analyser rapidement et fournir des projections transformables en action,» précisent les auteurs.

Ce rapport tombe à point nommé, puisqu'en 2015, «61% des décideurs américains et européens sur les données et l'analytique affirment que leur entreprise a déployé ou déjà un déploiement Big Data avant fin 2016,» rappelle le cabinet.

Selon Forrester, les entreprises clientes sont plus éclairées et exigeantes en matière de Big Data. En 2016, elles réclament plus d'intégration, ainsi que des APIs (bibliothèques de fonctions) afin d'intégrer le potentiel Big data dans leurs applications. Autre forte demande : des outils analytiques évolués (recourant au In-Memory), des solutions de préparation des données, ou plus d'analyse de données en self-service (voir illustration).

Afin de déterminer l'importance des diverses technologies et les nouvelles attentes du marché, Forrester Research a interrogé de multiples acteurs ou clients du Big Data à travers le monde : 63 fournisseurs, 17 experts, et 65 entreprises actuellement ou potentiellement utilisatrices. Bref, un panorama de tout l'écosystème actuel et à venir pour un instantané riche en informations et offrant une perspective lisible.

Deux turbos pour accélérer le Big Data accessible

Qualifiées de "technologies en phase de création", l'intelligence artificielle et le machine learning côtoient tout juste le Big Data. Objectif : répondre aux demandes des métiers avec l'auto-apprentissage, et limiter les compétences nécessaires à l'utilisation du Big data.

En arrière-plan du Big Data, le retour de **l'intelligence artificielle** va au-delà du machine learning. Elle propose le deep learning et des services sémantiques et de recherche statistique favorisant la mise en exergue du sens des informations et sa formalisation en langage compréhensible, avec plus de pertinence. Pour Forrester, IBM ne fait plus cavalier seul (avec Watson), et «plusieurs technologies et algorithmes voient le jour chaque mois.» Cependant, l'adoption reste lente (voire anecdotique), car les compétences sont encore peu répandues.

En parallèle, **la science des données** a ravivé l'intérêt pour le machine learning. Ici, divers algorithmes et moteurs de règles profitent de l'ajout continu de données pour détecter des schémas récurrents ou des "comportements", dans le but d'affiner les actions à mener ou d'interagir avec l'utilisateur. Là encore, ce domaine reste celui des data-scientists, aussi bien au fait des modèles statistiques et autres algorithmes prédictifs que des schémas de données. Des profils encore rarissimes, même si titre fleurit de façon ridicule dans nombre d'entreprises....

Du haut de ces plateformes... des pétaoctets vous contemplant

Dans la catégorie des “technologies pour la phase de survie”, affichant un plus haut degré de maturité et adoptées par de plus en plus d’entreprises, on retrouve 9 technologies. Elles sont déjà toutes adossées à des écosystèmes d’entreprises et de fournisseurs.

Trois d’entre elles ont déjà subi une évolution sensible, avec un succès parfois très relatif malgré leur pertinence pour l’entreprise. **Le Big Data as a Service** a évolué de “Hadoop as a Service” vers des plateformes Big Data, à même de simplifier l’accès à Hadoop (ou similaire) et de déployer rapidement ce type d’environnement. Forrester prédit un succès mesuré pour ces technologies, car peu d’entreprises devraient remplacer leurs piles logicielles Big Data, et les vendeurs ont privilégié une approche sur site plutôt qu’hybride. Cependant, tout un écosystème se développe pour les proposer en mode cloud ou hybride. A voir...

Plus que des règles, des politiques et des processus, **la gouvernance et l’audit de données** visent à favoriser la confiance dans les données, tout en intégrant les règles de conformité et la sécurité. Ces deux derniers aspects confortent Forrester dans sa prévision de forte croissance de ces solutions dans un proche avenir.

Comment appréhender le Big Data sans **la conception des données et la gestion des métadonnées**? Ingestion, accès, écritures... les cas d’usages des projets Open Source Apache Big Data (en lecture et écriture) gagnent de plus en plus en maturité avec -par exemple- l’autodécouverte de schémas de données (même complexes). Malgré leur grand intérêt, Forrester relativise le succès de ces outils, car les entreprises maintiennent leur processus manuels et visent plutôt ce qui apporte plus de valeur ajoutée immédiate aux métiers. Pourtant, qui veut voyager loin ménage sa monture!

La compréhension de la donnée devient une science

Mettre les données ou plutôt le sens caché et précieux des données à la portée de l’utilisateur métier reste LA promesse du Big Data. Pour y parvenir, les informaticiens et autres data analysts ou data scientists recourent de plus en plus aux logiciels de **préparation des données**, intégrant eux aussi des technologies comme Hadoop, Spark, etc. Sans surprise, ces technologies en pleine explosion sont promises à un bel avenir pour démocratiser enfin l’accès aux données. Incontournable dans la trousse “technologies pour la phase de survie”.

Aux côtés de la préparation, le travail sur **la qualité des données** est devenu indispensable au Big data: nettoyage (cleansing), transformation, enrichissement... et le tout le plus automatisé possible. Avec l’adoption de plus en plus importante du Big Data à l’échelle de toute l’entreprise, les auteurs du rapport misent sur une forte croissance de ces solutions, intégrant des innovations comme le machine learning.

Longtemps réservés aux statisticiens et autres super spécialistes de la donnée, **les outils de Data Science** se développent et se démocratisent sur fond d’Open Source. Par exemple, les bibliothèques de

machine learning, les codes de transformation, ou les outils de visualisation semblent plus accessibles. Selon les fournisseurs, ces outils sont à la portée des novices. Néanmoins, leur utilisation apparemment simple peut conduire à des résultats trompeurs... Forrester envisage d'ailleurs un succès modéré.

Des plateformes de détection de tendances (Insight platforms) sont devenues monnaie courante grâce aux packages offrant (en ligne ou sur site) des solutions intégrant stockage, récupération de données, Hadoop, fonctions analytiques, recherche simplifiée, etc. le tout en interaction ou intégré à des applications. En outre, ces outils utilisent la gestion et la sécurisation des données, voire le machine learning. Souvent issue de l'Open Source, ces plateformes devaient continuer à enregistrer un succès croissant.

Rendre les données et l'analytique accessibles à tous dans l'entreprise (selon ses attributions, évidemment) favorise la création de valeur pour les métiers. Une démarche qui avait déjà été initiée avec le langage SQL devenu l'espéranto de la data. Le succès du **SQL for Hadoop** (généralement open source) semblait donc facile à prédire. Forrester rapporte avoir déjà identifié 14 technologies SQL for Hadoop en 2015.

Ultime lame du couteau suisse des "technologies pour la phase de survie" selon Forrester, **l'ingestion de streams** (différente du Streaming Analytics, abordé plus loin) consiste à absorber en temps réel d'énormes flux continus de données : régulation du flux, intégration, opérations... Ces outils temps réel proviennent le plus souvent de projets open source. Indispensable pour juguler le tsunami de données, Forrester estime pourtant qu'il faudra attendre entre un et trois ans avant que la Stream Ingestion explose.

A lire aussi :

[BluData : comment Auchan bâtit son bras armé Big Data](#)

[Big Data : le Machine Learning protège les Livebox Orange de la foudre](#)

[Big Data : Mappy accélère son cluster Hadoop... sans acheter de serveurs](#)